APML - Image Denoising

יובל יעקבי, 302247077

7 בנובמבר 2018

חלק תאורטי,

שאלה 1.1 - MLE באלגוריתם הEM

בכיתה ראינו: $\mathbb{E}\left[l\left(S,\Theta\right)\right]=\sum_{i=1}^{N}\sum_{y=1}^{k}c_{i,y}log\left(\pi_{y}\mathcal{N}\left(x_{i};\mu_{y},\Sigma_{y}\right)\right)$

הראה כי הMLE של המשקולות שלנו היא אכן: $\pi_{y}=\frac{1}{N}\sum_{i=1}^{N}c_{i,y}$

__הוכחה__:

נזכור כי $\sum_{y=1}^{k}\pi_{y}=1$ מכיוון שכל פרמטר כזה מספר לנו על ה"הסיכוי" של הדגימה שלנו להיות מהגאוסיאן הy, כלומר הסכום של כולם צריך להיות 1.

נוכל להשתמש בכופלי לגראנז' על מנת למצוא את המקסימום של התוחלת

נשים לב כי

$$\begin{aligned}
\mathcal{L}\left(\pi_{1},\ldots,\pi_{k},\lambda\right) & =\sum_{i=1}^{N}\sum_{y=1}^{k}c_{i,y}log\left(\pi_{y}\mathcal{N}\left(x_{i};\mu_{y},\Sigma_{y}\right)\right)-\lambda\left(\sum_{y=1}^{k}\pi_{y}-1\right)=\\
 & =\sum_{y=1}^{k}\sum_{i=1}^{N}c_{i,y}\left(log\left(\pi_{y}\right)+log\left(\mathcal{N}\left(x_{i};\mu_{y},\Sigma_{y}\right)\right)\right)-\lambda\sum_{y=1}^{k}\pi_{y}+\lambda=\\
 & =\sum_{y=1}^{k}\sum_{i=1}^{N}c_{i,y}log\left(\pi_{y}\right)+\sum_{y=1}^{k}\sum_{i=1}^{N}c_{i,y}log\left(\mathcal{N}\left(x_{i};\mu_{y},\Sigma_{y}\right)\right)-\lambda\sum_{y=1}^{k}\pi_{y}+\lambda=\\
 & =\sum_{y=1}^{k}log\left(\pi_{y}\right)\sum_{i=1}^{N}c_{i,y}-\lambda\sum_{y=1}^{k}\pi_{y}+\sum_{y=1}^{k}\sum_{i=1}^{N}c_{i,y}log\left(\mathcal{N}\left(x_{i};\mu_{y},\Sigma_{y}\right)\right)+\lambda
\end{aligned}$$

כעת נגזור לי π_{y} (כלומר נקבל k משוואות) שנראות מהצורה הזו: (עבור $1\leq y\leq k$)

$$\frac{\partial}{\partial\pi_{y}}=\frac{\sum_{i=1}^{N}c_{i,y}}{\pi_{y}}-\lambda$$

נסביר את הגזירה - בכל גזירה i כל הרכיבים π_{y} עבור $y\neq i$ מתאפסים, ולכן הכפל גם מתאפס.

כעת אנחנו מחפשים את המקסימום ולכן נשווה את הנגזרת החלקית ל0

$$\frac{\sum_{i=1}^{N}c_{i,y}}{\pi_{y}}-\lambda=0\iff\frac{\sum_{i=1}^{N}c_{i,y}}{\lambda}=\pi_{y}$$

נגזור לפי λ ונשווה ל0

$$\frac{\partial}{\partial\lambda}=\sum_{y=1}^{k}\pi_y=1$$

נחבר את 3 המשוואות ונקבל את הנדרש

$$\sum_{y=1}^{k}\frac{\sum_{i=1}^{N}c_{i,y}}{\lambda}=1\rightarrow\sum_{i=1}^{N}\sum_{y=1}^{k}c_{i,y}=\lambda\xrightarrow{*}N\sum_{i=1}^{N}c_{i,y}=\lambda$$

* נשים לב כי $c_{i,y}$ מתאר את ההסתברות שתמונה i מגיעה מגאוסיאן y, לכן אם אנחנו סוכמים על כל העמודה ההסתברות היא 1, כיוון שזהו מרחב הסתברות תיקני.

שאלה 1.2 - MLE במודל GSM

צ"ל: שכלל העדכון של r_y אכן ימקסם את $r_y^2=\frac{\sum_{i=1}^{n}c_{i,y}x_i^T\Sigma^{-1}x_i}{d\sum_{i=1}^{n}c_{i,y}}$

<u>הוכחה</u>:

נזכור כי - MLE של התפלגות נורמלית הוא:

$$\mathbb{E}\left[l\left(S,\Theta\right)\right]=\sum_{i=1}^{N}\sum_{y=1}^{k}c_{i,y}log\left(\pi_y\mathcal{N}\left(x_i;0,r_y^2\Sigma_y\right)\right)$$

נפתח את הביטוי נגזור לפי r_y ונקווה לקבל את כלל העדכון הנכון:

$$\sum_{i=1}^{N}\sum_{y=1}^{k}c_{i,y}log\left(\pi_y\mathcal{N}\left(x_i;\mu_y,r_y\Sigma_y\right)\right)=\sum_{i=1}^{N}\sum_{y=1}^{k}c_{i,y}log\left(\pi_y\frac{1}{\sqrt{|2\pi r_y\Sigma|}}e^{-\frac{1}{2}x_i^T\left(r_y^2\Sigma\right)^{-1}x_i}\right)=$$
$$=\sum_{i=1}^{N}\sum_{y=1}^{k}c_{i,y}log\left(\pi_y\frac{1}{\sqrt{\left(2\pi r_y^2\right)^d det(\Sigma)}}\right)+\sum_{i=1}^{N}\sum_{y=1}^{k}c_{i,y}log\left(e^{-\frac{1}{2}x_i^T\left(r_y^2\Sigma\right)^{-1}x_i}\right)=$$
$$=\sum_{i=1}^{N}\sum_{y=1}^{k}c_{i,y}log\left(\pi_y\frac{r_y^{-d}}{\sqrt{\left(2\pi\right)^n det(\Sigma)}}\right)+\sum_{i=1}^{N}\sum_{y=1}^{k}c_{i,y}-\frac{1}{2}x_i^Tr_y^{-2}\Sigma^{-1}x_i=$$
$$=\sum_{i=1}^{N}\sum_{y=1}^{k}c_{i,y}log\left(\pi_y\frac{1}{\sqrt{\left(2\pi\right)^d det(\Sigma)}}\right)-\sum_{i=1}^{N}\sum_{y=1}^{k}c_{i,y}dlog\left(r_y\right)+\sum_{i=1}^{N}\sum_{y=1}^{k}\left(-\frac{1}{2}c_{i,y}x_i^Tr_y^{-2}\Sigma^{-1}x_i\right)$$

ע"פ התוכנית כעת נגזור לפי r_y:

$$\frac{\partial}{\partial r_y}=\sum_{i=1}^{N}\sum_{y=1}^{k}\frac{c_{i,y}x_i^T\Sigma^{-1}x_i}{r_y^3}-d\sum_{i=1}^{N}\sum_{y=1}^{k}\frac{c_{i,y}}{r_y}=\sum_{i=1}^{N}\sum_{y=1}^{k}\frac{c_{i,y}x_i^T\Sigma^{-1}x_i}{r_y^3}-\frac{dc_{i,y}}{r_y}$$
$$=\sum_{i=1}^{N}\sum_{y=1}^{k}\frac{c_{i,y}\cdot x_i^T\Sigma^{-1}x_i-r_y^2dc_{i,y}}{r_y^3}$$

ונשווה לאפס

$$\sum_{i=1}^{N}\sum_{y=1}^{k}\frac{c_{i,y}\cdot x_i^T\Sigma^{-1}x_i - r_y^2 d c_{i,y}}{r_y^3} = 0 \iff$$

$$\sum_{i=1}^{N}\sum_{y=1}^{k}\frac{c_{i,y}\cdot x_i^T\Sigma^{-1}x_i}{d\cdot c_{i,y}} = r_y^2 \overset{*}{\iff}$$

$$\sum_{i=1}^{N}\frac{c_{i,y}\cdot x_i^T\Sigma^{-1}x_i}{d\cdot c_{i,y}} = r_y^2 \iff$$

$$r_y^2 = \frac{c_{i,y}\sum_{i=1}^{N} x_i^T\Sigma^{-1}x_i}{d\sum_{i=1}^{N} c_{i,y}}$$

* כפי שהראנו בסעיף הקודם עמודות $c_{i,y}$ הן 1 ולכן לא משפיעות על הכפל כנדרש

שאלה 1.3 - איתחול של EM

נניח שהאתחול של המשתנים הוא כזה: $\forall y : \pi_y = \frac{1}{k}, \mu_y = \mu, \Sigma_y = \Sigma$
נתבונן באיטרציות הראשונות של האלגוריתם שלנו:

.1

$$c_{i,y} := \frac{\pi_y \mathcal{N}\left(x_i;\mu_y,\Sigma_y\right)}{\sum_{l=1}^{k}\pi_l\mathcal{N}\left(x_i;\mu_y,\Sigma_y\right)} = \frac{\frac{1}{k}\cdot\mathcal{N}\left(x_i;\mu_y,\Sigma_y\right)}{\sum_{l=1}^{k}\frac{1}{k}\mathcal{N}\left(x_i;\mu_y,\Sigma_y\right)} = \frac{1}{k}$$

$$\pi_y := \frac{1}{N}\sum_{i=1}^{N}c_{i,y} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{k} = \frac{1}{k}$$

$$\mu_y := \frac{\sum_{i=1}^{N}c_{i,y}\cdot x_i}{\sum_{i=1}^{N}c_{i,y}} = \frac{\sum_{i=1}^{N}\frac{1}{k}\cdot x_i}{\frac{N}{k}} = \frac{1}{N}\sum_{i=1}^{n}x_i = \mu_y$$

$$\Sigma_y := \frac{\sum_{i=1}^{N}c_{i,y}\left(x_i-\mu_y\right)\left(x_i-\mu_y\right)^T}{\sum_{i=1}^{N}c_{i,y}} = \Sigma_y$$

.2

$$c_{i,y} := \frac{\pi_y \mathcal{N}\left(x_i;\mu_y,\Sigma_y\right)}{\sum_{l=1}^{k}\pi_l \mathcal{N}\left(x_i;\mu_y,\Sigma_y\right)} = \frac{\frac{1}{k}\cdot \mathcal{N}\left(x_i;\mu_y,\Sigma_y\right)}{\sum_{l=1}^{k}\frac{1}{k}\mathcal{N}\left(x_i;\mu_y,\Sigma_y\right)} = \frac{1}{k}$$

$$\pi_y := \frac{1}{N}\sum_{i=1}^{N} c_{i,y} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{k} = \frac{1}{k}$$

$$\mu_y := \frac{\sum_{i=1}^{N} c_{i,y}\cdot x_i}{\sum_{i=1}^{N} c_{i,y}} = \frac{\sum_{i=1}^{N}\frac{1}{k}\cdot x_i}{\frac{N}{k}} = \frac{1}{N}\sum_{i=1}^{n} x_i = \mu_y$$

$$\Sigma_y := \frac{\sum_{i=1}^{N} c_{i,y}\left(x_i-\mu_y\right)\left(x_i-\mu_y\right)^T}{\sum_{i=1}^{N} c_{i,y}} = \Sigma_y$$

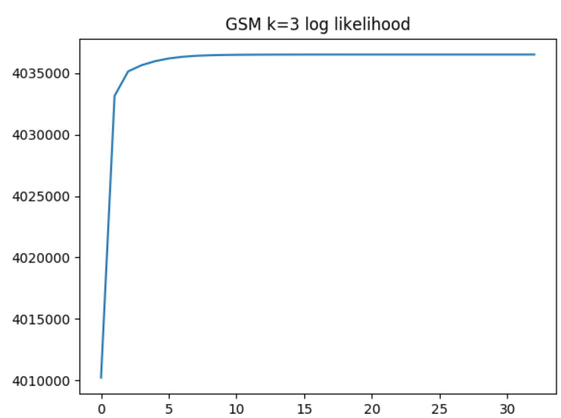
כפי שניתן לראות 2 האטרציות הן כבר אותו דבר ולכן גם בעתיד זה יראה ככה...

מדוע זה קורה? הגאוסיאנים נראים אותו הדבר, ולכן כל חלק תמונה שנראה לא נוטה להעדיף גאוסיאן כזה או אחר, ולכן פשוט נשאר עם אותם π שהתחלנו איתם שהם כאמור הסתברות אחידה...

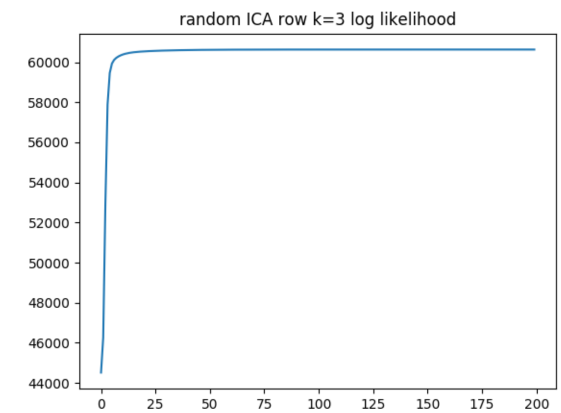
חלק פרקטי

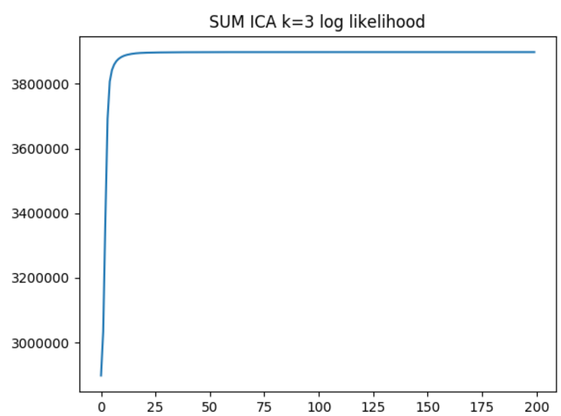
אלגוריתם EM

1. מודל GSM כאשר $k=3$ (מספר הגאוסיאנים)



2. מודל ICA כאשר $k=3$ (מספר הגאוסיאנים)





השוואות מודלים

זמן אימון - בשניות

- MVN - 0.0111
- GSM - 7.95
- ICA - 75

זמן ניקוי (denoise) של תמונה בודדת שניות:

- MVN - 23.2
- GSM - 104
- ICA - 192

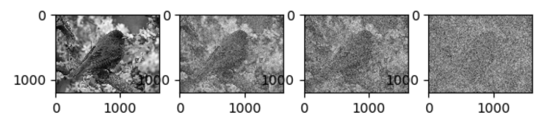
לוג likelihood על האימון

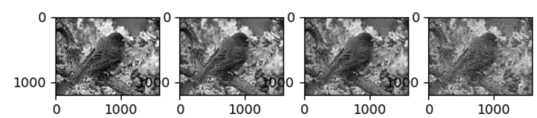
- MVN - $2.885 \cdot 10^6$
- GSM - $4.032 \cdot 10^6$
- ICA - $3.918 \cdot 10^6$

תמונות

כל התמונות מציגות רעש של: $0.01, 0.1, 0.2, 0.6$, והן התמונה הראשונה מתוך הtest_set

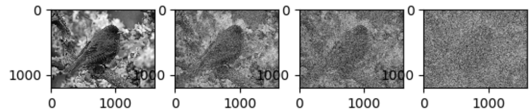
- MVN

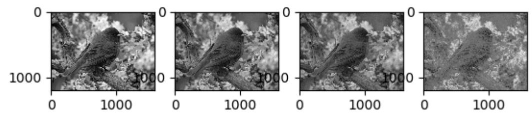




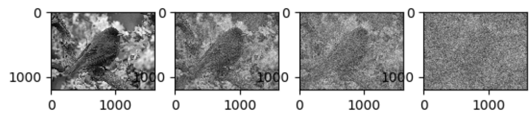
  - שגיאת MSE:
    - עבור רעש 0.01
      - תמונה רועשת - 10^{-4}
      - תמונה נקייה - $7.2 \cdot 10^{-5}$
      - שיפור - 72%
    - עבור רעש 0.1
      - תמונה רועשת - 10^{-3}
      - תמונה נקייה - 10^{-4}
      - שיפור - 10%
    - עבור רעש 0.2
      - תמונה רועשת - $4 \cdot 10^{-2}$
      - תמונה נקייה - $1.7 \cdot 10^{-3}$
      - שיפור - 4.5%
    - עבור רעש 0.6
      - תמונה רועשת - $36 \cdot 10^{-2}$
      - תמונה נקייה - $6.5 \cdot 10^{-3}$
      - שיפור - 1.8%

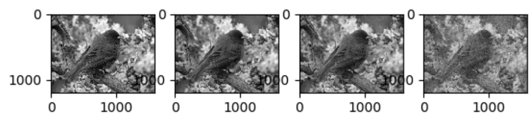
- GSM





  - שגיאת MSE:
    - עבור רעש 0.01
      - תמונה רועשת - 10^{-4}
      - תמונה נקייה - $9.3 \cdot 10^{-5}$
      - שיפור - 93%
    - עבור רעש 0.1
      - תמונה רועשת - 10^{-3}
      - תמונה נקייה - $11 \cdot 10^{-4}$
      - שיפור - 11%
    - עבור רעש 0.2
      - תמונה רועשת - $4 \cdot 10^{-2}$
      - תמונה נקייה - $2 \cdot 10^{-3}$
      - שיפור - 5%
    - עבור רעש 0.6
      - תמונה רועשת - $36 \cdot 10^{-2}$
      - תמונה נקייה - $9 \cdot 10^{-3}$
      - שיפור - 3%
- ICA





- שגיאת MSE:
  - עבור רעש 0.01
    - תמונה רועשת - 10^{-4}
    - תמונה נקייה - 10^{-4}
    - שיפור - 0%
  - עבור רעש 0.1
    - תמונה רועשת - 10^{-2}
    - תמונה נקייה - 10^{-3}
    - שיפור - 10%
  - עבור רעש 0.2
    - תמונה רועשת - $4 \cdot 10^{-2}$
    - תמונה נקייה - $1.7 \cdot 10^{-3}$
    - שיפור - 4.5%
  - עבור רעש 0.6
    - תמונה רועשת - $36 \cdot 10^{-2}$
    - תמונה נקייה - $6.6 \cdot 10^{-3}$
    - שיפור - 2%

סיכום מודלים - נראה שמודל GSM הוא הטוב ביותר

- מבחינת מהירות אימון - לא הכי מהיר, MVN כמובן מהיר יותר, אך אנו רואים כי ההתכנסות יחסית מהירה וזמן האימון לא משמעותי
- מבחינת שגיאה - השגיאה שלו הנמוכה ביותר
- כמו כן הנראות המרבית על הדאטא של האימון היא הגבוה ביותר - כלומר הוא לא מופתע לראות את הדאטא הזה, כלומר הוא מכליל כמו שצריך (: